

# Common variants conferring risk of schizophrenia

Hreinn Stefansson<sup>1\*</sup>, PhD, Roel A. Ophoff<sup>2,3\*</sup>, PhD, Stacy Steinberg<sup>1\*</sup>, PhD, Ole A. Andreassen<sup>4</sup>, MD PhD, Sven Cichon<sup>5</sup>, PhD, Dan Rujescu<sup>6</sup>, MD, Thomas Werge<sup>7</sup>, PhD, Olli P. H. Pietiläinen<sup>8,9</sup>, B.Sc, Ole Mors<sup>10</sup>, MD PhD, Preben B. Mortensen<sup>11</sup>, MD D.Med.Sc, Engilbert Sigurdsson<sup>12, 13</sup>, MD M.Sc, Omar Gustafsson<sup>1</sup>, PhD, Mette Nyegaard<sup>14</sup>, PhD, Annamari Tuulio-Henriksson<sup>15</sup>, PhD, Andres Ingason<sup>1</sup>, B.Sc, Thomas Hansen<sup>7</sup>, MSc, Jaana Suvisaari<sup>15</sup>, MD PhD, Jouko Lonnqvist<sup>15</sup>, MD PhD, Tiina Paunio<sup>16</sup>, MD PhD, Anders D. Børghlum<sup>10,14</sup>, MD PhD, Annette Hartmann<sup>6</sup>, PhD, Anders Fink-Jensen<sup>17</sup>, MD D.Med.Sc, Merete Nordentoft<sup>18</sup>, MD PhD DrMedSc, David Hougaard<sup>19</sup>, MD DrMedSc, Bent Norgaard-Pedersen<sup>19</sup>, MD DrMedSc, Yvonne Böttcher<sup>1</sup>, PhD, Jes Olesen<sup>20</sup>, MD PhD, René Breuer<sup>21</sup>, M.Sc, Hans-Jürgen Möller<sup>22</sup>, MD, Ina Giegling<sup>6</sup>, PhD, Henrik B. Rasmussen<sup>7</sup>, DVM PhD, Sally Timm<sup>23</sup>, MD, Manuel Mattheisen<sup>5</sup>, MD, István Bitter<sup>24</sup>, MD PhD, János M. Réthelyi<sup>24</sup>, MD PhD, Brynja B. Magnúsdóttir<sup>12,13</sup>, PhD, Thordur Sigmundsson<sup>12,13</sup>, MD, Pall Olason<sup>1</sup>, PhD, Gisli, Masson<sup>1</sup>, PhD, Jeffrey R. Gulcher<sup>1</sup>, MD PhD, Magnus Haraldsson<sup>12,13</sup>, MD, Ragnheidur Fossdal<sup>1</sup>, B.Sc, Thorgeir E. Thorgeirsson<sup>1</sup>, PhD, Unnur Thorsteinsdóttir<sup>1,13</sup>, PhD, Mirella Ruggeri<sup>25</sup>, MD PhD, Sarah Tosato<sup>25</sup>, MD, , MD, Barbara Franke<sup>26</sup>, PhD, Eric Strengman<sup>2</sup>, M.Sc, Lambertus A. Kiemeny<sup>27</sup>, PhD, GROUP, Ingrid Melle<sup>4</sup>, MD PhD, Srdjan Djurovic<sup>28</sup>, PhD, Lilia Abramova<sup>29</sup>, PhD DrSci, Vasily Kaleda<sup>29</sup>, PhD, Julio Sanjuan<sup>30</sup>, MD PhD, Rosa de Frutos<sup>31</sup>, PhD, Elvira Bramon<sup>32</sup>, MRCPsych PhD, Evangelos Vassos<sup>32, 33</sup>, MD PhD, Gillian Fraser<sup>34</sup>, BA, Ulrich Ettinger<sup>32,33</sup>, PhD, Marco Picchioni<sup>32</sup>, MRCPsych, Nicholas Walker<sup>35</sup>, MD, Timi Touloupoulou<sup>33</sup>, PhD, Anna C. Need<sup>36</sup>, PhD, Dongliang Ge<sup>36</sup>, PhD, Joeng Lim Yoon<sup>37</sup>, M.Sc, Kevin V. Shianna<sup>36</sup>, PhD, Nelson B. Freimer<sup>3</sup>, MD, Rita M. Cantor<sup>3, 37</sup>, PhD, Robin Murray<sup>32,33</sup>, FRCPsych DSc, Augustine Kong<sup>1</sup>, PhD, Vera Golimbet<sup>29</sup>, PhD DrSci, Angel Carracedo<sup>38</sup>, MD PhD, Celso Arango<sup>39</sup>, MD PhD, Javier Costas<sup>40</sup>, PhD, Erik G. Jönsson<sup>41</sup>, MD PhD, Lars Terenius<sup>41</sup>, PhD, Ingrid Agartz<sup>41</sup>, MD PhD, Hannes Petursson<sup>12,13</sup>, MD PhD, Markus M. Nöthen<sup>5</sup>, MD, Marcella Rietschel<sup>21</sup>, MD PhD, Paul M. Matthews<sup>42</sup>, MA (Oxon) MD DPhil FRCP, Pierandrea Muglia<sup>43</sup>, Leena Peltonen<sup>8,9</sup>, MD PhD, David St Clair<sup>34</sup>, MD PhD, David B. Goldstein<sup>36</sup>, MD PhD, Kari Stefansson<sup>1,13\*\*</sup>, MD PhD, David A. Collier, PhD<sup>32,44</sup>.

\*These authors contributed equally to this work

\*\*Corresponding author Dr Kari Stefansson, kari.stefansson@decode.is

- <sup>1</sup>CNS Division, deCODE genetics, Sturlugata 8, IS-101, Reykjavik, Iceland.
- <sup>2</sup>Department of Medical Genetics and Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG, Utrecht, The Netherlands.
- <sup>3</sup>UCLA Center for Neurobehavioral Genetics, Charles E. Young Drive South, Los Angeles, California 90024, USA.
- <sup>4</sup>Dept. of Psychiatry, Ullevål University Hospital and Institute of Psychiatry, University of Oslo, Kirkeveien 166, N-0407 Oslo, Norway.
- <sup>5</sup>Department of Genomics, Life & Brain Center, University of Bonn, Sigmund-Freud-Strasse 25, D-53127 Bonn, Germany.
- <sup>6</sup>Division of Molecular and Clinical Neurobiology, Department of Psychiatry, Ludwig-Maximilians-University, Nußbaumstrasse 7, 80336 Munich, Germany.
- <sup>7</sup>Research Institute of Biological Psychiatry, Mental Health Centre Sct. Hans Copenhagen University Hospital, DK-4000 Roskilde, Denmark.
- <sup>8</sup>Institute of Molecular Medicine, Biomedicum Helsinki, Haartmaninkatu 8, 00290 Helsinki, Finland.
- <sup>9</sup>Wellcome Trust Sanger Institute, Hinxton, UK.
- <sup>10</sup>Centre for Psychiatric Research, Aarhus University Hospital, Risskov, Skovagervej 2, 8240 Risskov, Denmark.
- <sup>11</sup>National Centre for Register-based Research, Aarhus University, Taasingegade 1, DK-8000 Aarhus, Denmark.
- <sup>12</sup>Department of Psychiatry, National University Hospital, Hringbraut, 101 Reykjavik, Iceland.
- <sup>13</sup>University of Iceland, School of Medicine, Laeknagardi, 101 Reykjavik, Iceland
- <sup>14</sup>Department of Human Genetics, The Bartholin Building, Aarhus University, DK-8000 Aarhus C, Denmark.
- <sup>15</sup>Department of Mental Health and Alcohol Research, National Public Health Institute, Mannerheimintie 166, FIN-00300 Helsinki, Finland.
- <sup>16</sup>Department for Molecular Medicine, National Public Health Institute, Biomedicum, Haartmaninkatu 8, 00290 Helsinki, Finland.
- <sup>17</sup>Mental Health Centre Rigshospitalet, Copenhagen University Hospital, DK-2100 Copenhagen Ø, Denmark.
- <sup>18</sup>Psychiatric Centre Bispebjerg, Building 13A, Bispebjerg Hospital, Bispebjerg Bakke 23, 2400 Copenhagen NV, Denmark.
- <sup>19</sup>Section of Neonatal Screening and Hormones, Dept. Clinical Chemistry and Immunology, The State Serum Institute, Artillerivej 5, 2300 Copenhagen S, Denmark.
- <sup>20</sup>Department of Neurology, 57 Nordre Ringvej, Glostrup Hospital, Glostrup, DK-2600 Copenhagen, Denmark.
- <sup>21</sup>Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, University of Heidelberg, J5, D-68159 Mannheim, Germany.

- <sup>22</sup>Department of Psychiatry, Ludwig-Maximilians-University, Nußbaumstrasse 7, 80336 Munich, Germany.
- <sup>23</sup>Mental Health Centre Frederiksberg, Copenhagen University Hospital, DK-2000 Frederiksberg, Denmark.
- <sup>24</sup>Semmelweis University, Department of Psychiatry and Psychotherapy, Budapest, Hungary.
- <sup>25</sup>Section of Psychiatry and Clinical Psychology, University of Verona, Verona, 37134 Verona, Italy.
- <sup>26</sup>Department of Human Genetics, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands.
- <sup>27</sup>Department of Epidemiology & Biostatistics and Department of Urology, Radboud University Nijmegen Medical Centre, PO Box 9101, 6500 HB Nijmegen, The Netherlands.
- <sup>28</sup>Department of Medical Genetics, Ulleval University Hospital and Institute of Psychiatry, University of Oslo, Kirkeveien 166, N-0407 Oslo, Norway.
- <sup>29</sup>Mental Health Research Center, Russian Academy of Medical Sciences, Zagorodnoe sh. 2/2, 117152 Moscow, Russia.
- <sup>30</sup>Unidad de Psiquiatría, Facultad de Medicina, Universidad de Valencia, CIBERSAM, Spain.
- <sup>31</sup>Departamento de Genética. Facultad de Biología. Universidad de Valencia, CIBERSAM, Spain.
- <sup>32</sup>Division of Psychological Medicine, Institute of Psychiatry, King's College, London SE5 8AF, UK.
- <sup>33</sup>Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College, London SE5 8AF, UK.
- <sup>34</sup>Department of Mental Health, University of Aberdeen, Royal Cornhill Hospital, Aberdeen AB25 2ZD, UK.
- <sup>35</sup>Ravenscraig hospital, Inverkip Road, Greenock PA16 9HA, UK.
- <sup>36</sup>Institute for Genome Sciences & Policy, Center for Population Genomics & Pharmacogenetics, 4011 GSRB II 103 Research Drive, Duke University, DUMC Box 3471, Durham, North Carolina 27708, USA.
- <sup>37</sup>Department of Human Genetics, UCLA, 695 Charles Young Drive South, Los Angeles, California 90095, USA.
- <sup>38</sup>Fundación Pública Galega de Medicina Xenómica-Complexo Universitario Hospitalario de Santiago, and CIBER de Enfermedades Raras (CIBERER), IML- Universidade de Santiago de Compostela, Spain.
- <sup>39</sup>Hospital General Universitario Gregorio Marañón, Centro de Investigación Biomédica en Red de Salud Mental, CIBERSAM, Madrid, Spain.
- <sup>40</sup>Fundación Pública Galega de Medicina Xenómica, and CIBER de Enfermedades Raras (CIBERER), Santiago de Compostela, Spain.
- <sup>41</sup>Department of Clinical Neuroscience, HUBIN project, Karolinska Institutet and Hospital, R5:00, SE-171 76 Stockholm, Sweden.
- <sup>42</sup>Clinical Imaging Centre, Clinical Pharmacology and Discovery Medicine, GlaxoSmithKline, Hammersmith Hospital, London W12 ONN, UK.
- <sup>43</sup>Medical Genetics, GlaxoSmithKline R&D, Via A. Fleming 4, 37135 Verona, Italy.

<sup>44</sup>Psychiatric Laboratory, Department of Psychiatry, West China Hospital, Sichuan University, Sichuan, China.

<sup>45</sup>Department of Psychiatry, Rudolf Magnus Institute of Neuroscience, University Medical Center Utrecht, The Netherlands.

<sup>46</sup>Academic Medical Centre University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands.

<sup>47</sup>Maastricht University Medical Centre, South Limburg Mental Health Research and Teaching Network, Maastricht, The Netherlands.

<sup>48</sup>University Medical Center Groningen, Department of Psychiatry, University of Groningen, The Netherlands.

## GROUP

René S. Kahn<sup>45</sup>, MD PhD, Don H. Linszen<sup>46</sup>, MD PhD, Jim van Os<sup>47</sup>, MD PhD, Durk Wiersma<sup>48</sup>, MD PhD, Richard Bruggeman<sup>48</sup>, MD PhD, Wiepke Cahn<sup>45</sup>, MD PhD, Lieuwe de Haan<sup>46</sup>, MD PhD, Lydia Krabbendam<sup>47</sup>, MD PhD, Inez Myin-Germeys<sup>47</sup>, MD PhD.

Schizophrenia is a complex disorder, caused by both genetic and environmental factors and their interactions. Research on pathogenesis has traditionally focused on neurotransmitter systems in the brain, particularly those involving dopamine.

Schizophrenia has been considered a separate disease for over a century, but in the absence of clear biomarkers, diagnosis has historically been based on signs and symptoms. A fundamental message emerging from genome-wide association studies (GWASs) of copy number variations (CNVs) associated with the disease is that its genetic basis does not necessarily conform to classical nosological disease boundaries. Certain CNVs confer not only high relative risk of schizophrenia but also of other psychiatric disorders<sup>1-3</sup>. The structural variations associated with schizophrenia can involve several genes and the phenotypic syndromes, or the “genomic disorders“, have not yet been characterized<sup>4</sup>.

Single nucleotide polymorphism (SNP)-based GWASs with the potential to implicate individual genes in complex diseases may reveal underlying biological pathways. In the present study we combined SNP data from several large genome-wide scans and followed up the most significant association signals. We found significant association with several markers spanning the MHC region on 6p22.1-22.3, a marker located upstream of the neurogranin gene (*NRGN*) on 11q24.2 and a marker in intron four of transcription factor 4 (*TCF4*) on 18q21.2. Our findings implicating the MHC region are consistent with an immune component to schizophrenia risk, while the association with *NRGN* and *TCF4* points to perturbation of pathways involved in brain development, memory and cognition.

To begin our search for sequence variants associated with schizophrenia, we performed a genome-wide scan of 2,663 schizophrenia cases and 13,498 controls from eight European locations (England, Finland/Helsinki, Finland/Kuusamo, Germany/Bonn, Germany/Munich, Iceland, Italy and Scotland; collectively called SGENE-plus) using the Illumina HumanHap300 and HumanHap550 BeadChips. In total, 314,868 SNPs meeting our quality control criteria were included in an allelic association analysis. To adjust for relatedness and potential population stratification, genomic control was used.

None of the markers gave  $P$  values smaller than our genome-wide significance threshold of  $0.05/314,868$ , or approximately  $1.6 \times 10^{-7}$  (see **Supplementary Figure 1** for a quantile-quantile plot and **Supplementary Table 1** for markers with the smallest  $P$  values). Next, we combined findings from our top 1,500 markers with results for these same markers (or surrogates for them) from both the International Schizophrenia Consortium<sup>5</sup> (ISC; excluding the Scottish samples overlapping with samples in our study, 2,602 cases / 2,885 controls) and the European-American portion of the Molecular Genetics of Schizophrenia<sup>6</sup> (MGS; 2,687 cases / 2,656 controls) study. Twenty-five of our top 1,500 markers (or 18 counting very strongly correlated ( $r^2 > 0.8$ ) markers only once) had  $P$  values less than  $1 \times 10^{-5}$  in the combined results (**Supplementary Table 2**). These top markers were followed up in as many as 4,999 cases and 15,555 controls from four sets of additional samples from Europe (set 1: 715 cases and 3,634 controls from the Netherlands; set 2: 3,330 cases and 6,892 controls from Denmark/Aarhus, Denmark/Copenhagen, Germany/Bonn, Germany/Munich, Hungary, the Netherlands, Norway, Russia and Sweden; set 3: 287 cases and 3,987

controls from Finland; set 4: 667 cases and 1,042 controls from Spain/Santiago and Spain/Valencia) (**Supplementary Table 3**).

Three markers, all in the extended major histocompatibility complex (MHC) region on the short arm of chromosome 6, were genome-wide significant in the combination of SGENE-plus and the follow-up samples described above (**Table 1**). In addition, four other markers, two in the MHC region, one at 11q24.2 and one at 18q21.2, were genome-wide significant when results from ISC and MGS were included (**Table 1**).

In the MHC region on chromosome 6p22.1-22.3, the five genome-wide significant markers ( $P$  ranging from  $1.1 \times 10^{-9}$  to  $1.4 \times 10^{-12}$  in all samples combined) have risk alleles with average control frequencies between 78% and 92% (**Table 1**). Combined odds ratios (ORs) for the markers range from 1.15 to 1.24 (**Table 1**) with no significant heterogeneity between the study groups ( $P > 0.25$ , **Supplementary Table 4**). For all of the markers, the multiplicative model for risk provides an adequate fit ( $P > 0.62$ ).

Despite spanning about five megabases (Mb), the five chromosome 6p markers cover only about 1.4 centimorgans (cM) and substantial linkage disequilibrium (LD) exists between them (**Supplementary Table 5**). The association of rs6932590 (the most significant marker), however, cannot account for all of the association of the four remaining markers (**Supplementary Table 6**). Most notably, conditional on rs6932590, rs3131296 has an association  $P$  value of  $3.4 \times 10^{-6}$ , suggesting that rs3131296 may be capturing a second susceptibility variant or that both rs6932590 and rs3131296 are correlated with a third, higher risk, variant not examined here.



To examine association of the genome-wide significant SNPs in the 5 Mb region on 6p22.1-22.3 with classical human leukocyte antigen (HLA) alleles, long-range phasing haplotypes<sup>7</sup> tagging the major alleles at the *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1* loci in Icelanders were used. Only rs3131296 shows substantial ( $r^2 > 0.5$ ) correlation with any of the classical HLA alleles tested; this marker has an  $r^2$  of 0.86 with DRB1\*03 and an  $r^2$  of 0.81 with HLA-B\*08. Simplified tags for these two classical alleles, appropriate for the European samples of SGENE-plus, had effects that were not statistically distinguishable from the effect of rs3131296. In the case of both DRB1\*03 and HLA-B\*08, the classical HLA allele is paired with the protective allele of rs3131296, making the results described here consistent with the under-transmission of DRB1\*03 to schizophrenic offspring reported by Li et al.<sup>8</sup>

Many autoimmune and infectious diseases have been associated with DRB1\*03 and, indeed, inspection of top MHC region SNPs from recent genome-wide association scans of three of these, type I diabetes (T1D)<sup>9</sup>, celiac disease (CD)<sup>10</sup> and systemic lupus erythematosus<sup>11</sup>, reveals, for each disease, SNPs having a HapMap CEU  $r^2$  of at least 0.73 with rs3131296. For all of the diseases, the allele protective from schizophrenia is associated with the allele at-risk for the autoimmune disease. This reciprocal association may, at least in part, explain the recently reported inverse association between T1D and schizophrenia<sup>12,13</sup>. A positive association, however, has been described for CD and schizophrenia<sup>12</sup>; if this positive association has a genetic basis, it must be the result of associations at variants other than the one described here.

Schizophrenia patients are more likely, compared to the general population, to have been born in the winter or the spring. Although infections such as influenza and measles have been proposed as a possible mechanism for this distortion, a clear association between infectious agents and schizophrenia has not been demonstrated. The association with the MHC region reported here supports a role for infection but, as many non-immune related genes are also found in the extended MHC region, it does not provide strong evidence. Based on the 3,130 schizophrenia patients for which month of birth information was available, no significant difference in the frequency of the top SNPs from the MHC region according to season of birth (winter/spring versus summer/fall) was identified ( $P > 0.29$ ).

The MHC region has long been postulated to harbor variants conferring risk of schizophrenia, both because of evidence for linkage in the region<sup>14</sup> and because of the suggested involvement of infection. Association studies of variants from the MHC region to date, however, have had modest sample sizes and therefore have lacked power to detect effects similar to those described here.

The genome-wide significant marker ( $P = 2.4 \times 10^{-9}$ ) at 11q24.2, rs12807809, has an average risk allele control frequency of 83% and a combined OR of 1.15 (**Table 1**) with no significant OR heterogeneity between the study groups ( $P = 0.74$ , **Supplementary Table 4**). The multiplicative model provides an adequate fit ( $P = 0.18$ ). This marker is 3,457 bases upstream of neurogranin (*NRGN*). *NRGN* has previously been reported as associated in males with schizophrenia in a small Portuguese series<sup>15</sup> although the associated SNP in that paper, rs7113041, is not closely correlated with the

SNP reported here (HapMap CEU  $r^2=0.11$ ). Furthermore, reduced NRG1 immuno-reactivity has been observed in prefrontal areas 9 and 32 of postmortem schizophrenia brains<sup>16</sup>. *NRG1* is expressed exclusively in brain, especially in dendritic spines, with expression directly controlled by thyroid hormone<sup>17</sup>. It is therefore possible that the psychotic and cognitive features associated with thyroid dysfunction, may in part be mediated through dysregulation of *NRG1* gene expression.

*NRG1* encodes a postsynaptic protein kinase substrate that binds calmodulin (CaM) in the absence of calcium; it is abundantly expressed in brain regions important for cognitive functions, and is especially enriched in CA1 pyramidal neurons in the hippocampus<sup>18</sup>. The main function of NRG1 may be to act as a CaM reservoir, regulating its availability in the postsynaptic compartment. Glutamate stimulation of N-methyl-d-aspartate (NMDA) receptors results in  $Ca^{2+}$  influx to the neuron, NRG1 oxidation and release of CaM<sup>19</sup>. The consequent activation of postsynaptic calcium/calmodulin- dependent protein kinase II (CaMKII) by CaM results in a sustained strengthening of synaptic connections; conversely CaM activation of calcineurin (PP2B) weakens these connections. CaMKII plays a major role in mediating the NMDA-receptor signaling involved in synaptic plasticity and formation of associative memories in the brain<sup>20</sup>. Impaired memory function is thought to be a core feature of the pathophysiology of schizophrenia<sup>21</sup>, especially affecting short term memory where CaMKII is thought to play a major role<sup>22</sup>. Glutamate stimulation of NMDA receptors results in  $Ca^{2+}$  influx to the neuron and in NRG1 oxidation. Altered NRG1 activity may therefore mediate the effects of NMDA hypofunction implicated in the pathophysiology of schizophrenia.

On 18q21.2, marker rs9960767 has a genome-wide significant  $P$ -value of  $4.1 \times 10^{-9}$  (**Table 1**). The risk allele control frequency is about 6% and the OR is 1.23 with no significant OR heterogeneity between the study groups ( $P = 0.34$ , **Supplementary Table 4**). The multiplicative model gives an adequate fit ( $P = 0.81$ ). Genome scan meta analysis of linkage studies of schizophrenia<sup>14</sup> ranks the 18q21.1-qter 'bin' around 15<sup>th</sup> in the genome. *TCF4*, is essential for normal brain development<sup>23</sup> and mutations in the gene were recently found to be responsible for Pitt Hopkins syndrome (PHS), an autosomal dominant neurodevelopmental disorder characterized by severe motor and mental retardation, microcephaly, epilepsy and facial dysmorphisms<sup>24</sup>. The phenotype need not be as extreme as PHS, a *de novo* translocation disrupting exon 4 of *TCF4* was found in an individual with problems restricted to mental retardation alone<sup>25</sup>. Thus, it seems that variants in a single gene can be associated with a range of neuropsychiatric phenotypes including schizophrenia. This is in line with the range of phenotypes associated with CNVs recently associated with schizophrenia<sup>1-3</sup>.

In addition to these three genome-wide significant loci, further putative susceptibility variants are highlighted by this study. For instance, in the set of 18 markers taken into follow-up studies, rs2312147 achieved a  $P$  value that was not far from our genome-wide significance threshold (**Supplementary Table 3**). As well, markers having  $P$  values in the combined SGENE-plus, ISC and MGS data set (**Supplementary Table 2**) somewhat larger than the cutoff used here ( $1 \times 10^{-5}$ ) can be investigated further. Intriguingly, these markers include rs6589386, which is located in an intergenic region upstream of *DRD2*, a candidate gene for schizophrenia.

Our findings demonstrating association with markers in the MHC region are consistent with previous reports suggesting immune system involvement in schizophrenia, while association with *NRGN* and *TCF4* points more to perturbation of pathways involved in brain development and cognitive function, particularly memory. Impaired cognitive and memory functions are being recognized increasingly as core features of schizophrenia<sup>21</sup> which are poorly addressed by current medications. The three common genetic variants we describe, which predispose to schizophrenia, have the potential to be translated into targets for the development of novel medications.

## **METHODS SUMMARY**

**Subjects.** SGENE-plus, the genome-wide portion of the study, included 2,663 cases and 13,498 controls from eight European locations - England, Finland/Helsinki, Finland/ Kuusamo, Germany/Bonn, Germany/Munich, Iceland, Italy and Scotland. Follow-up groups comprised 4,999 cases and 15,555 controls from twelve European locations (Denmark/Aarhus, Denmark/Copenhagen, Finland, Germany/Bonn, Germany/Munich, Hungary, the Netherlands, Norway, Russia, Spain/Santiago, Spain/Valencia and Sweden). Cases were diagnosed with schizophrenia according to DSM-IV or ICD-10 criteria (see Supplementary Methods).

**Genotyping.** Genome-wide genotyping was carried out at deCODE, Duke University and the University of Bonn using either HumanHap300 or HumanHap550 BeadChips (Illumina). Individual genotyping was done via Centaurus assays at deCODE and multiplex PCR and

minisequencing assays, followed by MALDI-TOF mass spectrometry analysis in Spain and Finland.

**Statistical analysis.** A likelihood procedure described previously was used for association analysis<sup>26</sup>. To correct for relatedness and potential population stratification, genomic control was used<sup>27</sup>. Within our study, samples were combined with the Mantel-Haenszel model<sup>28</sup>. Our results were merged with those of the ISC and MGS by computing weighted averages of *Z* scores (see full Methods online).

## Acknowledgements

We want to thank the subjects and their relatives and staff at the recruitment centers. This work was sponsored by EU grants LSHM-CT-2006-037761 (Project SGENE), PIAP-GA-2008-218251 (Project PsychGene) and HEALTH-F2-2009-223423 (Project PsychCNVs). Genotyping of the Dutch samples was sponsored by NIMH funding, R01 MH078075. This work was also supported by the National Genomic Network (NGFN-2) of the German Federal Ministry of Education and Research (BMBF) and Marie Curie grant PIAP-GA-2008-218251(PsychGene). M.M.N. received support from the Alfried Krupp von Bohlen und Halbach-Stiftung. We are grateful to S. Schreiber and M. Krawczak for providing genotype data for PopGen controls, and to K.-H. Jöckel and R. Erbel for providing control individuals from the Heinz Nixdorf Recall Study. Recruitment of the patients from Munich was partially supported by GlaxoSmithKline (GSK). We are grateful to the Genetics Research Centre (GRC) GmbH, an initiative by GSK and LMU. The Northern Finland Birth Cohort 1966 (NFBC66) is thanked for providing population controls for the study. The genotyping of NFBC66 was financially supported by National Institutes of Health grant 1R01HL087679-01, STAMPEED.

**Author Contributions:** H.S., S.S., D.C., D.S.C., D.R., E.S. and K.S. wrote the first draft of the paper. M.H., B.B.M., K.D.J., P.M., I.G., H.-J.M., A.H., A.C.N., C.C., G.F., N.W., J.L., J.S., A.T., T.T., E.B., M.D.F., R.M., M.R., S.T., M.W., C.V., T.W.M., A.G.W., H.U., S.D., I.M., J.O., O.A.A., A.G., M.R., R.O., J.B., R.S.K., L.A.K., O.G., A.D.B., M.N., A.D.B., A.F.J., M.N., D.H., B.N., Y.B., R.B., H.B.R., S.T., M.M., I.B., J.R., I.S., L.A., V.K., J.S., R.F., E.V., U.E., M.P., J.L.Y., N.B.F., R.C., V.G., A.C., C.A., J.C., E.J., L.T., I.A., O.M., P.B.M., and B.F. recruited, diagnosed and gathered phenotypes. H.S., D.R., R.F., E.S., T.S., P.M., T.T., J.R.G., U.T., H.P., D.G., T.W., D.C., L.P., D.S.C. and K.S. planned, supervised and coordinated the work. H.S., S.C., P.O., G.M., A.I., S.S., A.G., T.E.T., O.P.H.P., B.V.H., D.G., K.V.S., M.M.N., T.H. and A.K. analyzed the data. All authors contributed to the current version of the paper.

## References

1. Stefansson, H. et al. Large recurrent microdeletions associated with schizophrenia. *Nature* **455**, 232-6 (2008).
2. Mefford, H. C. et al. Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N Engl J Med* **359**, 1685-99 (2008).
3. International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature* **455**, 237-41 (2008).
4. Brunetti-Pierri, N. et al. Recurrent reciprocal 1q21.1 deletions and duplications associated with microcephaly or macrocephaly and developmental and behavioral abnormalities. *Nat Genet* **40**, 1466-71 (2008).
5. International Schizophrenia Consortium. Common genetic variants play a large role in schizophrenia. *Nature* (Submitted).
6. Shi, J. & Levinson, D. F. Common variants on chromosomes 6p22.1 and 3q26.33 are associated with schizophrenia. *Nature* (Submitted).
7. Kong, A. et al. Detection of sharing by descent, long-range phasing and haplotype imputation. *Nat Genet* (2008).
8. Li, T. et al. Transmission disequilibrium analysis of HLA class II DRB1, DQA1, DQB1 and DPB1 polymorphisms in schizophrenia using family trios from a Han Chinese population. *Schizophr Res* **49**, 73-8 (2001).
9. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-78 (2007).
10. van Heel, D. A. et al. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* **39**, 827-9 (2007).
11. Harley, J. B. et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* **40**, 204-10 (2008).
12. Eaton, W. W. et al. Association of schizophrenia and autoimmune diseases: linkage of Danish national registers. *Am J Psychiatry* **163**, 521-8 (2006).
13. Juvonen, H. et al. Incidence of schizophrenia in a nationwide cohort of patients with type 1 diabetes mellitus. *Arch Gen Psychiatry* **64**, 894-9 (2007).
14. Lewis, C. M. et al. Genome scan meta-analysis of schizophrenia and bipolar disorder, part II: Schizophrenia. *Am J Hum Genet* **73**, 34-48 (2003).
15. Ruano, D. et al. Association of the gene encoding neurogranin with schizophrenia in males. *J Psychiatr Res* **42**, 125-33 (2008).
16. Broadbelt, K., Ramprasad, A. & Jones, L. B. Evidence of altered neurogranin immunoreactivity in areas 9 and 32 of schizophrenic prefrontal cortex. *Schizophr Res* **87**, 6-14 (2006).
17. Bernal, J., Rodriguez-Pena, A., Iniguez, M. A., Ibarrola, N. & Munoz, A. Influence of thyroid hormone on brain gene expression. *Acta Med Austriaca* **19 Suppl 1**, 32-5 (1992).
18. Huang, F. L., Huang, K. P. & Boucheron, C. Long-term enrichment enhances the cognitive behavior of the aging neurogranin null mice without affecting their hippocampal LTP. *Learn Mem* **14**, 512-9 (2007).
19. Li, J., Pak, J. H., Huang, F. L. & Huang, K. P. N-methyl-D-aspartate induces neurogranin/RC3 oxidation in rat brain slices. *J Biol Chem* **274**, 1294-300 (1999).
20. Bliss, T. V. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31-9 (1993).



21. Touloupoulou, T. et al. Substantial genetic overlap between neurocognition and schizophrenia: genetic modeling in twin samples. *Arch Gen Psychiatry* **64**, 1348-55 (2007).
22. Wang, H. et al. CaMKII activation state underlies synaptic labile phase of LTP and short-term memory formation. *Curr Biol* **18**, 1546-54 (2008).
23. Flora, A., Garcia, J. J., Thaller, C. & Zoghbi, H. Y. The E-protein Tcf4 interacts with Math1 to regulate differentiation of a specific subset of neuronal progenitors. *Proc Natl Acad Sci U S A* **104**, 15382-7 (2007).
24. Pitt, D. & Hopkins, I. A syndrome of mental retardation, wide mouth and intermittent overbreathing. *Aust Paediatr J* **14**, 182-4 (1978).
25. Kalscheuer, V. M. et al. Disruption of the TCF4 gene in a girl with mental retardation but without the classical Pitt-Hopkins syndrome. *Am J Med Genet A* **146A**, 2053-9 (2008).
26. Gretarsdottir, S. et al. The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. *Nat Genet* **35**, 131-8 (2003).
27. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997-1004 (1999).
28. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* **22**, 719-48 (1959).
29. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-59 (2000).

## METHODS

**Genome-wide genotyping.** SGENE was initially made up of 1,321 cases and 12,277 controls typed at deCODE Genetics using the Illumina HumanHap300 BeadChip. For SGENE-plus, an additional 859 cases and 854 controls typed at Duke University using either the Illumina HumanHap300 BeadChip or the Illumina HumanHap550 BeadChip as well as 483 cases and 367 controls typed at the University of Bonn using the HumanHap550 BeadChip were also included. Samples were excluded if they were low-yield (low yield was defined as below 98% except for the samples typed at Duke in which case low-yield was below 96%), if they were duplicates of other samples included in the study, if they had a sex determined by X chromosome marker homozygosity different from their reported sex or if they were estimated to have less than 90% European ancestry by running STRUCTURE<sup>29</sup> using the HapMap CEU, YRI and CHB/JPT individuals as reference samples. Of the 317,503 markers on the HumanHap300 BeadChip, 2,635 were deemed unusable due to lack of polymorphism, severe deviation from Hardy-Weinberg equilibrium ( $P < 1 \times 10^{-10}$ ), low yield ( $< 95\%$  in either cases or controls) or allele frequency differences between the typing centers ( $P < 1 \times 10^{-7}$ ) ; 314,868 markers, then, remained for analysis.

**Follow-up genotyping.** Follow-up set 1 (715 cases; 3,634 controls) was genotyped at UCLA on the HumanHap550 BeadChip and at deCODE Genetics on the HumanCNV370 BeadChip. Only the markers shown in Supplementary Table 3, however, were used in this study. Follow-up set 2 (3,330 cases; 6,892 controls) was genotyped at deCODE Genetics using Centaurus assays (Nanogen). Assay quality was evaluated by genotyping the CEU HapMap samples and comparing the results with the publicly released HapMap data. Assays with a greater than 1.5% mismatch rate were not used. Follow-up set 3 (287 cases; 3,987 controls) was typed in Finland using the Sequenom MassArray iPLEX genotyping system, following the manufacturer's

instructions (Sequenom Inc., San Diego, CA, USA). Briefly, the system involves multiplex PCR and mini-sequencing assays, followed by MALDI-TOF mass spectrometry analysis. Follow-up set 4 (667 cases; 1,042 controls) was typed at the Santiago de Compostela node of the Spanish National Genotyping Centre ([www.cegen.org](http://www.cegen.org)) using the Sequenom MassArray iPLEX genotyping system, following the manufacturer's instructions (Sequenom Inc., San Diego, CA, USA). As a quality check, all clusters were manually inspected for accurate genotype assignment. In addition, 1781 genotypes were successfully assayed twice, with no discordant results.

**Association analysis.** Association analysis was carried out using a likelihood procedure described in a previous publication implemented in NEMO software<sup>26</sup>. Allele-specific ORs and associated *P*-values were calculated assuming a multiplicative model for the two chromosomes of an individual. Association was tested using a standard likelihood ratio statistic that, if the subjects were unrelated, would have asymptotically a chi-square distribution with 1 degree of freedom under the null hypothesis. To correct for relatedness and potential population stratification in the genome-wide typed samples (SGENE-plus and follow-up set 1), genomic control was used<sup>27</sup>. Inflation factors, estimated by dividing the median of the 314,868 chi-square statistics by 0.675<sup>2</sup>, were 1.01, 1.03, 1.09, 1.05, 1.05, 1.19, 1.04 and 1.09 for the SGENE-plus England, Finland/Helsinki, Finland/Kuusamo, Germany/Bonn, Germany/Munich, Iceland, Italy and Scotland groups, respectively, and 1.08 for follow-up set 1. The inflation factor was large in Iceland because of the inclusion of close relatives in that data set. Both SGENE-plus and the follow-up samples were combined using the Mantel-Haenszel model<sup>28</sup>.

**Summary statistic combination.** Combined *P* values for the SGENE-plus, ISC and MGS studies were calculated by summing *Z* scores with each group's *Z* score multiplied by the inverse of that group's standard error divided by the square root of the sum of the squared

inverse standard errors. Combined ORs were calculated by summing log ORs with each log OR weighted by the inverse of its variance.

**Table 1. Genome-wide significant association of seven markers with schizophrenia**

Chr/Mb	SNP[allele]	Freq	SGENE-plus <sup>1</sup> (2,663 cases / 13,498 controls)		Follow-up (4,999 cases / 15,555 controls)		SGENE-plus + Follow-up (7,662 cases / 29,053 controls)		S-p + F-u + ISC + MGS (12,951 cases / 34,594 controls)		Region/ neighboring gene
			OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	
6/27.2	rs6913660[C] <sup>2,8</sup>	0.85	1.22 (1.10,1.36)	0.00023	1.11 (1.04,1.19)	0.0021	1.14 (1.08,1.21)	4.7x10 <sup>-6</sup>	1.15 (1.10,1.21)	1.1x10 <sup>-9</sup>	MHC/ <i>HIST1H2BJ</i>
6/27.3	rs13219354[T] <sup>3,8</sup>	0.90	1.25 (1.11,1.42)	0.00043	1.19 (1.08,1.30)	0.00022	1.21 (1.12,1.30)	4.4x10 <sup>-7</sup>	1.20 (1.14,1.27)	1.3x10 <sup>-10</sup>	MHC/ <i>PRSS16</i>
6/27.4	rs6932590[T] <sup>4,8</sup>	0.78	1.15 (1.05,1.26)	0.0024	1.17 (1.10,1.25)	4.9x10 <sup>-7</sup>	1.17 (1.11,1.23)	4.4x10 <sup>-9</sup>	1.16 (1.11,1.21)	1.4x10 <sup>-12</sup>	MHC/ <i>PRSS16</i>
6/28.4	rs13211507[T] <sup>5,8</sup>	0.92	1.24 (1.08,1.42)	0.0027	1.27 (1.15,1.40)	3.1x10 <sup>-6</sup>	1.26 (1.16,1.36)	3.1x10 <sup>-8</sup>	1.24 (1.16,1.32)	8.3x10 <sup>-11</sup>	MHC/ <i>PGBD1</i>
6/32.3	rs3131296[G] <sup>5,8</sup>	0.87	1.21 (1.08,1.36)	0.0011	1.20 (1.11,1.30)	5.3x10 <sup>-6</sup>	1.21 (1.13,1.29)	2.1x10 <sup>-8</sup>	1.19 (1.13,1.25)	2.3x10 <sup>-10</sup>	MHC/ <i>NOTCH4</i>
11/124.1	rs12807809[T]	0.83	1.19 (1.08,1.32)	0.00045	1.13 (1.06,1.21)	0.00022	1.15 (1.09,1.22)	5.0x10 <sup>-7</sup>	1.15 (1.10,1.20)	2.4x10 <sup>-9</sup>	<i>NRGN</i>
18/51.3	rs9960767[C] <sup>6,8</sup>	0.056	1.30 (1.11,1.51)	0.0011	1.20 (1.08,1.33)	0.00044	1.23 (1.13,1.34)	2.2x10 <sup>-6</sup>	1.23 (1.15,1.32)	4.1x10 <sup>-9</sup>	<i>TCF4</i>

Allelic OR and *P* values (two-sided) based on the multiplicative model are shown. Frequency is the average allelic control frequency in SGENE-plus. Mb is from the National Center for Biotechnology Information (NCBI) build 36. To combine the study groups within SGENE-plus and within the follow-up, the Mantel-Haenszel model was used. SGENE-plus and the follow-up sets were combined with ISC and MGS by summing weighted summary statistics (the Aberdeen samples were not included in the ISC results used). <sup>1</sup>*P* values were adjusted using genomic control (see Methods) <sup>2</sup>rs4452638 in ISC (HapMap CEU *r*<sup>2</sup>=0.866) <sup>3</sup>rs4452638 in ISC (HapMap CEU *r*<sup>2</sup>=1) <sup>4</sup>rs3800307 in ISC (HapMap CEU *r*<sup>2</sup>=0.843) <sup>5</sup>rs13214023 in ISC (HapMap CEU *r*<sup>2</sup>=0.915) <sup>6</sup>rs1150753 in ISC (HapMap CEU *r*<sup>2</sup>=1) <sup>7</sup>rs10401120 in ISC (HapMap CEU *r*<sup>2</sup>=0.867) <sup>8</sup>imputed using MACH 1.0 in MGS

**Table 1. Genome-wide significant association of seven markers with schizophrenia**

Chr/Mb	SNP[allele]	Freq	SGENE-plus <sup>1</sup> (2,663 cases / 13,498 controls)		Follow-up (4,999 cases / 15,555 controls)		SGENE-plus + Follow-up (7,662 cases / 29,053 controls)		S-p + F-u + ISC + MGS (12,951 cases / 34,594 controls)		Region/ neighboring gene
			OR	P value	OR	P value	OR	P value	OR	P value	
6/27.2	rs6913660[C] <sup>2,8</sup>	0.85	1.22	0.00023	1.11	0.0021	1.14	4.7x10 <sup>-6</sup>	1.15	1.1x10 <sup>-9</sup>	MHC/ <i>HIST1H2BJ</i>
6/27.3	rs13219354[T] <sup>3,8</sup>	0.90	1.25	0.00043	1.19	0.00022	1.21	4.4x10 <sup>-7</sup>	1.20	1.3x10 <sup>-10</sup>	MHC/ <i>PRSS16</i>
6/27.4	rs6932590[T] <sup>4,8</sup>	0.78	1.15	0.0024	1.17	4.9x10 <sup>-7</sup>	1.17	4.4x10 <sup>-9</sup>	1.16	1.4x10 <sup>-12</sup>	MHC/ <i>PRSS16</i>
6/28.4	rs13211507[T] <sup>5,8</sup>	0.92	1.24	0.0027	1.27	3.1x10 <sup>-6</sup>	1.26	3.1x10 <sup>-8</sup>	1.24	8.3x10 <sup>-11</sup>	MHC/ <i>PGBD1</i>
6/32.3	rs3131296[G] <sup>5,8</sup>	0.87	1.21	0.0011	1.20	5.3x10 <sup>-6</sup>	1.21	2.1x10 <sup>-8</sup>	1.19	2.3x10 <sup>-10</sup>	MHC/ <i>NOTCH4</i>
11/124.1	rs12807809[T]	0.83	1.19	0.00045	1.13	0.00022	1.15	5.0x10 <sup>-7</sup>	1.15	2.4x10 <sup>-9</sup>	<i>NRGN</i>
18/51.3	rs9960767[C] <sup>7,8</sup>	0.056	1.30	0.0011	1.20	0.00044	1.23	2.2x10 <sup>-6</sup>	1.23	4.1x10 <sup>-9</sup>	<i>TCF4</i>

Allelic OR and *P* values (two-sided) based on the multiplicative model are shown. Frequency is the average allelic control frequency in SGENE-plus. Mb is from the National Center for Biotechnology Information (NCBI) build 36. To combine the study groups within SGENE-plus and within the follow-up, the Mantel-Haenszel model was used. SGENE-plus and the follow-up sets were combined with ISC and MGS by summing weighted summary statistics (the Aberdeen samples were not included in the ISC results used). <sup>1</sup>*P* values were adjusted using genomic control (see Methods) <sup>2</sup>rs4452638 in ISC (HapMap CEU *r*<sup>2</sup>=0.866) <sup>3</sup>rs4452638 in ISC (HapMap CEU *r*<sup>2</sup>=1) <sup>4</sup>rs3800307 in ISC (HapMap CEU *r*<sup>2</sup>=0.843) <sup>5</sup>rs13214023 in ISC (HapMap CEU *r*<sup>2</sup>=0.915) <sup>6</sup>rs1150753 in ISC (HapMap CEU *r*<sup>2</sup>=1) <sup>7</sup>rs10401120 in ISC (HapMap CEU *r*<sup>2</sup>=0.867) <sup>8</sup>imputed using MACH 1.0 in MGS